

Language Evolution and Information Theory

Rudolf Ahlswede¹ Erdal Arikan² Lars Baumer¹ Christian Deppe¹

¹Dept. of Mathematics
University of Bielefeld
Postfach 100131
D-33501 Bielefeld

² Department of Electrical and
Electronics Engineering
Bilkent University, TR-06800
Ankara, Turkey

Abstract — We study Nowak’s model for language evolution and settle a conjecture by him.

The human language is used to store and transmit information. Therefore there is significant interest in the mathematical models of language development. These models aim to explain how natural selection can lead to the gradual emergence of human language. Nowak and coworkers created such a mathematical model [2], [3]. A language \mathcal{L} in Nowak’s model is a system $\mathcal{L} = (\mathcal{O}, \mathcal{X}^n, d, r)$ consisting of the following elements

- \mathcal{O} is a finite set of objects, $\mathcal{O} = \{o_1, \dots, o_N\}$.
- \mathcal{X} is a finite set of phonemes which model the elementary sounds in the spoken language. The set \mathcal{X}^n models the set of all possible words of length n .
- Each object is mapped to a word by the function $r : \mathcal{O} \rightarrow \mathcal{X}^n$. Thus, the words for all objects have the same length n . The model allows several objects to be mapped to the same word. With some abuse of notation, we use \mathcal{L} to denote the set of all words in the language, $\mathcal{L} = \{x^n : x^n = r(o_i) \text{ for some } 1 \leq i \leq N\}$.
- $d : \mathcal{X} \times \mathcal{X} \rightarrow R_+$ is a measure of distance between phonemes; i.e., a function that is symmetric $d(x, y) = d(y, x)$ and non-negative $d(x, y) \geq 0$, with $d(x, y) = 0$ if and only if $x = y$. The distance between two words is defined by $d_n(x^n, y^n) = \sum_{i=1}^n d(x_i, y_i)$, where $x^n, y^n \in \mathcal{X}^n$, $x^n = (x_1, \dots, x_n)$, $y^n = (y_1, \dots, y_n)$.
- The model postulates that the conditional probability of the event that the listener understands the word $y^n \in \mathcal{L}$ given that the speaker utters the word $x^n \in \mathcal{L}$ is given by

$$p(y^n | x^n) = \frac{\exp(-d_n(x^n, y^n))}{\sum_{v^n \in \mathcal{L}} \exp(-d_n(x^n, v^n))}$$

Nowak defined the fitness of a language \mathcal{L} with words over \mathcal{X}^n as

$$F(\mathcal{L}, \mathcal{X}^n) = \sum_{x^n \in \mathcal{L}} p(x^n | x^n)$$

Nowak was interested in the maximum possible fitness for languages. So, he defined the fitness of the space \mathcal{X}^n as

$$F(\mathcal{X}^n) = \sup\{F(\mathcal{L}, \mathcal{X}^n) : \mathcal{L} \text{ is a language over } \mathcal{X}^n\}$$

and he posed the determination of the quantity $F(\mathcal{X}^n)$ for general spaces (X, d) as an open problem. He conjectured that $F(\mathcal{X}^n) = (F(\mathcal{X}))^n$ when (\mathcal{X}, d) is a metric space, i.e., when the distance function d satisfies the triangle inequality $d(x, y) + d(y, z) \geq d(x, z)$. We show that Nowak’s conjecture is true for a class of spaces defined by a certain condition on the distance function. Let us call a space (X, d) a p.s.d. space if the matrix $[e^{-d(x, y)}]_{x \in \mathcal{X}, y \in \mathcal{X}}$ is positive semi-definite. The main result is the following

Theorem 1 For any p.s.d. space (\mathcal{X}, d) where \mathcal{X} is a finite set, the fitness is given by

$$F(\mathcal{X}^n) = F(\mathcal{X})^n = e^{nR_0} \quad (1)$$

where

$$R_0 = R_0(\mathcal{X}, d) = -\log \min_{\lambda} \sum_x \sum_y \lambda_x \lambda_y e^{-d(x, y)} \quad (2)$$

where the minimum is over all probability distributions $\lambda = (\lambda_1, \dots, \lambda_{|\mathcal{X}|})$ on \mathcal{X} .

In other words, for p.s.d. spaces Nowak’s conjecture holds and the fitness is given by powers of e^{R_0} . For any p.s.d. space, there exists a “channel” $[W(z|x)]_{x \in \mathcal{X}, z \in \mathcal{Z}}$ for some set \mathcal{Z} such that (i) $W(z|x) \geq 0$, all x, z , (ii) $\sum_z W(z|x) = 1$, all x , and (iii) $e^{-d(x, y)} = \sum_z \sqrt{W(z|x)W(z|y)}$, all x, y . The parameter R_0 equals the cutoff rate of the channel W in the standard information-theoretic sense. This indicates a connection between Nowak’s model and standard information-theoretic models. Indeed, the proof of the above result makes use of Gallager’s results on reliability exponents and specifically his “parallel channels theorem” [1, p. 149] to achieve the single-letterization demanded by Nowak’s conjecture. Examples of spaces (\mathcal{X}, d) for which Nowak’s conjecture is settled by the above result are (i) the Hamming space where \mathcal{X} is an arbitrary finite set and $d(x, y) = \delta_{x, y}$ is the Hamming metric, (ii) \mathcal{X} is a finite set of reals and $d(x, y) = |x - y|$, and (iii) \mathcal{X} is a finite set of reals and $d(x, y) = (x - y)^2$. All of these spaces are p.s.d. Some other partial results are as follows: (i) All finite ultra-metric spaces are p.s.d. (Recall that in an ultra-metric space for all three points a, b, c it holds that $d(a, b) \leq \max\{d(a, c), d(c, b)\}$.) (ii) All metric spaces with 3 and 4 elements are p.s.d. (iii) There exists some metric spaces with 5 elements which are not p.s.d. (iv) For every metric space (\mathcal{X}, d) where \mathcal{X} is a subset of reals, there exists a scaling $d_\alpha(x, y) = \alpha d(x, y)$ for some $\alpha > 0$ and for all $x, y \in \mathcal{X}$ such that the space (\mathcal{X}, d_α) is p.s.d. (v) Nowak’s conjecture does not hold if we do not allow multiplicity of words.

ACKNOWLEDGMENTS

We would like to thank V. Blinovskiy and E. Telatar for discussions on this problem.

REFERENCES

- [1] R.G. Gallager, *Information Theory and Reliable Communication*. New York : Wiley, 1968.
- [2] M.A. Nowak and D.C. Krakauer, “The evolution of language”, *PNAS* 96, 14, 8028-8033, 1999.
- [3] M.A. Nowak, D.C. Krakauer and A. Dress, “An error limit for the evolution of language”, *Proceedings of the Royal Society Biological Sciences Series B*, 266, 1433, 2131-2136, 1999.